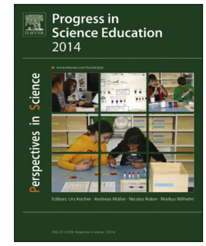
Perspectives in Science (2014) **2**, 5–21

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/pisc



Context-based science education by newspaper story problems: A study on motivation and learning effects ☆

Jochen Kuhn[a,*], **Andreas Müller**[b]

[a]*University of Kaiserslautern, Department of Physics/Physics Education Group, Erwin-Schrödinger-Str. 46, 67663 Kaiserslautern, Germany*
[b]*University of Geneva, Faculty of Sciences/Physics Department and Institute of Teacher Education, Pavillon d'Uni Mail (IUFE), Boulevard du Pont d'Arve 40, 1211 Genève, Switzerland*

Received 16 March 2014; accepted 1 June 2014
Available online 21 June 2014

KEYWORDS
Context-based science education;
Newspaper story problems;
Authenticity;
Narrative embedding;
Anchored instruction

Abstract
Using real-life contexts has a long-standing tradition and is considered as an important issue in both current science education and educational psychology. Drawing on research in both areas, as well as on practice reports from science classrooms, the present contribution deals with a specific form of establishing such contexts, viz. context by newspaper story problems (NSP). Using the particular form of science problems based on newspaper articles and the real-life contexts provided by them, effects on both motivation and learning were studied.
In a quasi-experimental comparison of 6 physics classes of secondary level 1 ($N=122$; grade 10, topic: energy) learning with newspaper based problems vs. conventional textbook problems (same content, lesson plan and teacher) showed considerable positive effects. This holds for general motivation, including several subscales ($p<0.01$, $\omega^2=0.52$) as well as for achievement, including transfer ($p<0.01$, $\omega^2=0.20$). Moreover, these results show robustness towards various individual and classroom features (e.g. gender, non-verbal intelligence and school type), and at least mid-term temporal stability. Newspaper story problems thus appear as a useful element of context-based science teaching.

Introduction: context-based science education and story contexts

Several recent reviews point out that context-based approaches and real-life connections are currently considered as a central issue in science education in general (Fensham, 2009; Bennett et al., 2007) and in physics

☆This article is part of a special issue entitled "Progress in Science Education 2014".

*Corresponding author. Tel.: +49 631 205 2393.
E-mail addresses: kuhn@physik.uni-kl.de (J. Kuhn), Andreas.Mueller@unige.ch (A. Müller).

http://dx.doi.org/10.1016/j.pisc.2014.06.001

education in particular (Taasoobshirazi and Carr, 2008; Kuhn, 2005, 2010; Kuhn and Müller, 2005a, 2005b). In a broad understanding of the term, context based science education (CBSE) is defined as "using concepts and process skills in real-life contexts that are relevant to students from diverse backgrounds" (Glynn and Koballa, 2005, p. 75). Making (or trying to do so) science issues relevant to students themselves, their families and their peers is opposed to the wide-spread perception of especially physics (or more generally: science) as being dry, impersonal and irrelevant, and this is supposed to have positive effects both on motivation and learning (Bennett et al., 2007). PISA (OECD, 2006) follows a similar understanding of CBSE, repeatedly emphasizing the importance of tasks and problems "that could be part of the actual experience or practice of the participant in some real-world setting", and it "places most value on tasks that could be encountered in a variety of real-world situations" (as can be seen also from the very items used in the study). Moreover, PISA points out the following feature of context-based learning: problems encountered in real-world settings are usually not stated in the disciplinary terms to be learned or applied. Thus, a kind of "translation", i.e. a terminological and conceptual reframing is initiated, representing an important step of cognitive activation. It has to be emphasized that even such a basic understanding of "context" is far from being trivial or educationally shallow, even though there is a discussion around the idea of CBSE with positions going far beyond that (see e.g. Gilbert et al., 2011). The present contribution pursues this line of research and development and aims at combining the general approach of CBSE with a specific format of establishing contexts, viz. "*stories* as context". Beginning, embedding, and connecting teaching content and sequences with an interesting story is a promising way of relating it to contexts beyond school. A particular form for this are newspaper story problems (NSP). These are problems related to newspaper articles containing science related issues, and which are (up to minor modifications) unchanged in both text and layout (see Fig. 1a).

From a practical point of view, the double rationale behind NSP is that (i) newspapers and newspaper articles as such stand for out-of-school, real-life contexts per se and (ii) journalists are supposed to be experts for writing interesting, good stories (so it is good advice to draw on this know-how). Good practice reports about successful realizations and existing collections of examples of using newspapers for mathematics and science literacy purposes are available, both on the international and several national levels (extensively in mathematics, see e.g. Herget and Scholz, 1998; Paulos, 1995). The same is true to some extent in biology (Gardner et al., 2009; Hoots, 1993; Jarman and McClune, 2001) and in chemistry (Haupt, 2005; Glaser and Carson, 2005; Toby, 1997) as well as in physics education (Armbrust, 2001). Jarman and McClune (2007) give an excellent introduction with many examples about the use of newspapers in science education in general.

For a review on uses and purposes of science teaching with newspapers see Jarman and McClune (2002). Having CBSE in mind it is interesting to note that within their sample (in Northern Ireland) "links with everyday life" were by far those most frequently stated as main intention (76%) and main benefit (62%). From a theoretical point of view, Norris and Phillips (2003) have convincingly argued that literacy in the basic or fundamental sense (including newspapers) is central to scientific literacy.

Moreover, the idea has a long-standing tradition for general literacy purposes, from the "Use the News" series in the Journal of Reading (Kossack, 1987) to the "Newspapers in Education (NIE)" programmes of several national newspaper associations (KMK/BDZV, 2006; NAAF, 2007, 2010a, 2010b, 2011).

From the research point of view, Fensham (2009) emphasizes in the "implications for teaching" section of his review the opportunities of a "context by story" approach to science teaching, in particular for motivational/affective aspects. As one important research finding among others, he states that the unexpected gender neutrality found for many countries in PISA 2000 "to have resulted, at least in

a **'Newspaper story problem' format**	b **'Traditional task' format**
Lausanne/ngn. Another great round-a-world adventure a la Steve Fossett's solo flight last month. Now it seems the explorer Bertrand Piccard will attempt the world's first solar powered around the world flight Piccard comes from a family of explorers and made history in March of 1999 with a nonstop, around-the-world flight in a hot-air balloon, the Breitling Orbiter 3. But what about the solar power side of this? Is that really possible? Nearly the entire body of the plane will be covered by 287 square yards (240 square metres) of solar panels. Piccard estimates that enough power can be generated to sustain a flight of roughly 60 miles an hour (97 kilometres an hour). The plane's batteries are going to have to be pretty heavy, capable of storing 200 watts per kilogram, so that the plane can run at night. *Gadling*, 2005-04-13	In 2007 the explorer Bertrand Piccard will attempt the world's first solar powered around the world flight. He estimates that the solar panels of the plane can generated enough power to sustain a flight of roughly 60 miles an hour (97 kilometres an hour). The plane's batteries are going to have to be pretty heavy, capable of storing 200 watts per kilogram, so that the plane can run at night.

1. How long will Bertrand Piccard be on his way around the world?
2. How much electrical energy per kilogram and how much power per kilogram has to be produced at least?
3. How much electrical energy per kilogram and how much power per kilogram has to be produced at least by the solar panels, if Piccard will drive ca. 75% of the flight by day?
4. How much electrical energy per kilogram has to be produced at least by the batteries?
5. Discuss your results critically, e.g. with respect to the energy transformation process, and use physical arguments thereto.

Fig. 1 "Newspaper story problem" format (a) and traditional task format (b). Questions are identical in both cases (lower part).

part, from a number of the presenting contexts being stories that involved people and science" (Fensham, 2009).

Looking more closely, research has put forward several theoretical and empirical arguments in favor of "context by story", "narrative contextualization" and similar approaches, and explanations of its potential. These will be reviewed in the following, both for motivation and cognition/learning.

Research background

Context by story

Motivation

Regarding motivation, an essential virtue of stories is the psychological (i.e. subjective) reality and familiarity human beings ascribe to them, from early age on (Mandler, 1987, 2004). Connecting curricular (e.g. scientific) content with a narrative context (e.g. through NSP) is supposed to transfer or "inherit" the familiarity of the latter to the former, thus helping to overcome the well-known cold and impersonal image of the sciences. While this is a clear and plausible argument, it has still to be established empirically, whether NSP (as a particular form of story based context) are really perceived as motivating by learners.

These general theoretical arguments on the "flavor of reality" of narrative contexts can be specified for teaching and learning based on newspapers in terms of several important aspects. Rhoades and Rhoades (1980) have drawn attention to *usefulness* as an important factor in the perception of newspapers. This is based on the experience that newspapers are a major source of information on a variety of issues of practical life, from serious (job, health etc.) to more pleasant questions (leisure, fashion, sports etc.). Again, the perception of usefulness is supposed to be transferred from newspapers to teaching and learning based on them. A further potentially important factor emphasized by Rhoades and Rhoades (1980) is fostering the student's *self-concept* as one important component of motivation (Shavelson et al., 1976; Hattie, 2009) by offering an opportunity of participation: newspapers enable young people to engage in conversations with adults (and peers), thereby opening up communication, the feeling and the experience of having something to say in various social contexts – a feature fostering a positive self-concept of probably anybody, not only of youths. If this turns out to be true, it would be educationally welcome, as a meta-analysis on science curriculum development performed by Shymansky et al. (1983) has shown that science-related self-concepts are usually hard to improve (positive effects were found on all 18 investigated outcomes *except for* self-concept). Furthermore Hattie (2009) stated that the hardest area to change was related to learned attributions (e.g., self-efficacy and effort). NSP as a vehicle to support such change would be a plausible and interesting hypothesis to be tested.

Finally, a general issue in the context of motivation has to be addressed. It is a widespread belief in education, that better motivation will lead to better learning, and this is also an explicit rationale behind a lot of work on CBSE (see e.g. Bennett et al., 2007). Generally, however, correlations between motivational and learning measures are lower than expected, generally around $r=0.30$ (Uguroglu and Walberg, 1979; Wild et al., 2001). For the PISA study in 2006, Fensham (2009) even discusses a weak negative correlation (OECD subsample, $r=-0.06$). With these general findings, it is necessary to consider arguments other than correlational supporting improved learning by NSP, which will be done in the next section.

Cognition and learning

Regarding cognition and learning, a first relevant and well-established research finding about narrative contexts is about improving memory for content. As it is stated e.g. in the entry on long-term memory of the *Encyclopedia of educational psychology* (Salkind, 2008, p. 620, and further literature cited there), "weaving the events to be remembered into a simple story or narrative is effective" (Salkind, 2008, p. 860; as a quantitative example for this narrative embedding, an improvement of accuracy for remembering world lists by a factor of 7 could be established).

Beyond this general finding about memory improvement, a specific cognitive process was established in the context of story memory, viz. their organization as "schemata" (Anderson, 2010). These are considered as "cognitive patterns of domain-specific information that are used as templates by individuals to help them explain, interpret, perceive, encode, and respond to complex tasks and experiences. [...] They create meaning from situations, data, and events by organizing and determining the patterns in complex sets of information" (Salkind, 2008, p. 864). An impressive line of research has established (Rumelhart, 1975; Mandler and Johnson, 1977; Mandler, 1984) that stories are perceived, organized and memorized as schemata in this sense,[1] and they are even seen as paradigmatic examples, as is supported by the following statement: "Probably the most powerful general schema that people anywhere possess is the knowledge of how stories are organized" (Salkind, 2008, p. 864).

Cognitive schemata in general and narrative schemata in particular support learning in at least two fundamentally important ways:

(i) by providing a cognitive pattern for the organization and interpretation of new experiences and of existing memory content, and
(ii) by enabling expectations about things which will or ought to happen, and similarly for reconstruction (which is expectation, shifted into the past).

Hence, stories and story schemata are offering an important way for the construction of meaning, and thus for meaningful learning (Zabel, 2004, 2007).

As a further point on cognition and learning a very important problem, common to most forms of context based learning, has to be addressed, viz. the question of transfer. In view of the well-known problem of "inert knowledge" (Renkl et al., 1996), it is not clear that, after having streng-

[1] According to current terminology, story or narrative schemata belong to the class of *script* schemata (Salkind, 2008).

thened the embedding into a context (which by definition is at the heart of CBSE), whether the *de*contextualisation necessary for transfer is easy to achieve. Transfer is especially important for the guiding idea of scientific literacy, which is behind much of the work on CBSE (Fensham, 2009; Roberts, 2007), and where linking to some context is not only a promising instructional approach, but a main purpose of education. With this tension of context and decontextualisation, possible difficulties with transfer are of particular concern. Indeed, PISA has found quite salient difficulties of this type (e.g. for German students; see Baumert et al., 2001). Thus, the question of transfer has to be kept in mind also when using story contexts.

On the theoretical and empirical basis explained above, we will now turn to a framework offering detailed design principles for classroom implementation of context-based science learning with newspaper story problems.

Anchored instruction and its design principles: motivation *and* learning

Probably one of the most developed instructional approaches to combine a perspective on motivation *and* learning on the one hand and narrative contexts as one key feature on the other, is „Anchored Instruction" (AI). It is one of the leading "schools" of Situated Learning (together with 'Cognitive Flexibility Theory', 'Cognitive Apprenticeship', and 'Goal-Based Scenario'; CTGV, 1990). AI has been developed since 1990 by the 'Cognition and Technology Group' in Vanderbilt (CTGV), led by J.D. Bransford (Bransford et al., 1990; Bransford and Stein, 1993; CTGV, 1990, 1991, 1992, 1993, 1997). It is it is distinguished (among the other schools of situated learning), and most close to the present research, by having the approach of narrative embedding (or contextualization) as one of its fundamental ideas. Moreover, it developed for this approach a specific form of instructional material ("anchor media") and researched based design principles, which can be transferred to the instructional tasks (NSP) investigated in this study. This will be discussed in the following paragraphs.

The basis of this approach is the conviction that teaching and learning should be anchored in realistic, motivating contexts, demanding the solution of authentic, meaningful problems. The central key to initiate this process are 'anchor media' (or 'anchors', for short). The original AI-approach uses interactive multimedia videodiscs, among which several series were developed for science education (e.g. the 'Jasper Woodbury Series'; see CTGV, 1997). Each series consists of different videodiscs, which lasts 15–20 min and have to be designed in a specific way described by the Vanderbilt group as follows (CTGV, 1993): "*The design of these anchors was quite different from the design of videos that were typically used in education…our goal was to create interesting, realistic contexts that encouraged the active construction of knowledge by learners. Our anchors were stories rather than lectures and were designed to be explored by students and teachers*". The anchor characteristics emphasized here and focused on in the present contribution are "active construction", authenticity ("realistic contexts") and a narrative, motivating embedding ("story character"). A particular strength of Anchored Instruction and its characteristics is the fact that its idea of situatedness combines fostering of both cognition and motivation: appropriate anchor problems can create meaningful contexts, where motivational and cognitive activation should go hand in hand.[2]

Indeed, the benefits of AI were shown in more than a dozen of studies, well summarized in the meta-analysis of Blumschein (2003). A weighted average of explained variance $\langle r^2 \rangle \approx 0.14$ was found[3] (corresponding to an effect size on the boundary from medium to large, see Cohen, 1988), with values up to $r^2=0.66$ (Bottge et al., 2002) for solving contextualized problems (a main purpose AI was invented for). Note, that AI thus offers considerable support for the theoretical expectation (explained in the preceding subsection on "Cognition and Learning"), that story contexts can foster meaningful learning. Moreover, it does so by using the "embedding" form of story contexts (mentioned above), where students are supposed to work and learn with various problems related to the embedded science content. AI has thus both sound theoretical and empirical support, and the NSP approach was strongly inspired by it. Concerning however a broader implementation of its idea, and their further development in classroom practice, there are some difficulties put forward in particular by both educational researchers and teachers interested in classroom innovation. A first difficulty with multimedia anchors is the considerable amount of time (and money) necessary for their development, usually far beyond the budgets available in schools.[4] Moreover, in most cases the necessary technological know-how cannot be assumed to be already present, which for broad classroom implementation requires even more unrealistic expenses for training. Two more difficulties teachers are particularly worried about, is the small flexibility of multimedia anchors with respect to curricular and instructional features, and the large extent to which a change of the teaching script is required by AI. A given classroom situation is defined by topics to be covered, length, complexity (and other features) appropriate for the particular class being taught and the like, all of which cannot be easily changed or adapted in videodiscs or other multimedia software (or only at the expense of the large investment of resources as already mentioned). Moreover,

[2] Actually, this can be seen as origin of the "anchor" metaphor of AI, which offers a kind of learning which is both motivationally and cognitively "anchored" in the specific form of contexts it is working with.

[3] Computed by Blumschein (2003, Table 37). Note that the weighted average of explained variance is *not* the square of the weighted average correlation coefficient, or mathematically $\langle r^2 \rangle \neq \langle r \rangle^2$.

[4] Quantitatively, the decisive factor is the ratio of time of development to time of instruction for a given medium. A lower estimate pertinent for the development phase of the Jasper series is 100, i.e. at least 100 h of development are necessary for one hour running time of the instruction medium, and this value rapidly increases with increasing functionality (e.g. degree of interactivity, reference possibility for background information) to 500 and more (Brahler et al., 1999). This is far beyond the time budget in schools, and with (very conservative) 35 USD per development hour also beyond the money budget (3500 USD per hour of instructional running time).

the very far-reaching change of the teaching script required by AI is very often not feasible (or desirable) for a given teacher in a given teaching situation. Thus, anchor media as they exist are meeting serious obstacles for a successful implementation in classroom and school reality.

Our study on NSP (and similar "anchor media", Kuhn, 2010; Kuhn and Müller, 2005a, 2005b; Müller et al., 2010) was inspired by AI and an attempt to overcome the difficulties of the original approach described above. While preserving authenticity, 'story'-character (narrative contexts) and student centered activity as design principles, it aims at an improved applicability to and implementation in a wider range of realistic educational settings, as text-based anchors are much easier and less expensive to develop and to modify than multimedia based anchors. The advantage of combining the general theoretical framework of narrative contexts, explained above, with design principles inspired by AI is that the latter already is based on a considerable body of evidence (see above) and has specific design principles to offer. Beyond those already mentioned, AI (and to a large extent also the present work) is also based on the following ones (CTGV, 1991)[5]:

Embedded data: the data necessary to solve a problem are "embedded" in the story of the learning anchor, and not given explicitly (as in conventional textbook problems). The rationale behind this design principle is as follows: (i) it is true for problems encountered in the real world (daily life, workplace, genuine research; cf. problem authenticity); (ii) the "translation" feature (OECD, 2006) is extended by a feature of "selection" of what is relevant from what is not (for a given problem), both contributing to cognitive activation. For these reasons, "embedded data" are considered as an especially important characteristic of AI.

Related problems (multiple contexts): learning should provide repeated opportunity and multiple contexts to acquire new concepts, not merely for the sake of repetition, but in order to avoid inert knowledge (cf. above); for single contexts, there is the danger of having the involved concepts "welded" to them (CTGV, 1991). The number of related problem stories (anchors) for the acquisition of new conceptual (and procedural) knowledge thus should be at least two (for the AI anchors) or more (for the shorter NSP anchors).

Collaborative learning: small group work, complemented by whole-class phases, ensures communication and social embedding considered necessary for active learning (social context or situatedness); this is also natural and easy to realize for the NSP approach (and actually a common element of contemporary science teaching in the authors' country).

Horizontal (cross-disciplinary) and vertical (cross-grade, cumulative learning) connections, which again help to strengthen the perception of relevant contexts and to overcome inert knowledge: these features also hold for newspaper story problems: horizontal links are included by construction, NSP involving links to many other issues, such as societal, technological, biological, etc.; vertical links are possible in principle and are taken into account by the specific problems posed for a given newspaper article.

Complex Problems and generative learning: for AI, presentation of the 15–20 min video stories, was followed by a rather open problem statement in form of a real-life goal. Students were supposed to find ("generate") themselves the intermediate, science (or other discipline) related questions to be solved for this goal, and the entire instructional setting (long story, embedded data, multistep problems, generative learning) leads to a rather high complexity. While this indeed is close to many real-life problem settings, it entails considerable, at a given learning level maybe hardly surmountable difficulties. In terms of instructional psychology, there is a dilemma of complexity vs. cognitive load, which cannot be decided a priori, but requires empirical investigation. For this purpose, complexity must be variable, necessitating a flexible, easy-to-change learning anchor (such as NSP). In the present study, problems were less complex than in AI (but still encompassing transfer and discussion, see the section on problem levels in "Materials and Methods"). This is close to current teaching practice, but the entire approach is also appropriate for studying *variable* degrees of complexity (Kuhn, 2007, 2008; Kuhn and Müller, 2006, 2007).

Summing up, the approach presented here is a form of CBSE based on work on narrative contexts and, regarding its design principles, more specifically inspired by Anchored Instruction. While most of the design features mentioned above are maintained, the video-based "anchors" of the original AI approach were of course deliberately replaced by newspaper story problems.

Research questions and hypotheses

In line with existing research described in the preceding section, the following research questions were examined, beginning with two questions on general effects: first, whether science learning with newspaper story problems is more motivating than learning with content-identical, conventional counterparts. Second, whether it is also more effective for learning, and to which degree.

Furthermore, whether these general beneficial effects also cover more specific aspects, closely connected to the theoretical background of the approach: third, whether perceived *self-concept* (as motivation sub-dimension) can be improved (because this is a feature of particular importance to CBSE in general, and NSP in particular). Fourth, whether the same is true for *transfer ability* (as learning sub-dimension, again essential for CBSE and, more generally, for scientific literacy).

Finally, there are two questions which are important for practical implementation (a main objective of the present study), viz. robustness (with respect cofactors) and stability (with respect to time): Fifth, to which extent beneficial effects – if existing – are robust with respect to pertinent learner and classroom characteristics (such as various academic achievement levels, gender, school type, and others). Beyond the practical issue, this is also a relevant conceptual aspect, as the CBSE in the understanding followed here should be "relevant to students from diverse backgrounds" (see Introduction and Glynn and Koballa, 2005). A question

[5]But beyond merely changing the medium, the present approach also differs from AI by the extent to which a change of the teaching script is implied.

of particular interest is whether the finding of gender independence (Fensham, 2009) can be replicated. Sixth, whether learners' motivation is temporally stable (at least at a mid-term range).[6]

Based on the theoretical framework explained above, our hypotheses are as follows: (1) Motivation after learning with newspaper story problems is higher than after learning with conventional text-book type problems. (2) Achievement after learning with newspaper story problems is higher than with conventional text-book type problems. Moreover, these general beneficial effects hold for (3) self-efficacy and (4) transfer ability in particular. (5) Finally, if there are positive effects, they should not be restricted to learners with specific features or background.

The remaining research question, viz. temporal stability, is important for both conceptual and practical reasons within the CBSE framework, but no specific hypothesis seems justifiable on theoretical grounds.

Materials and methods

Sample, study and teaching procedure

The above research questions and hypotheses were approached in a quasi-experimental design (motivation: pre-, post-, follow-up-test; achievement: post-test) with a NSP learning group (treatment group, TG) vs. conventional learning problem group (control group, CG) looking for motivation, learning (achievement) and possible interactions (see following sections) on the individual and classroom level. Note that the two groups were different in arrangement and layout of the instructional material (newspaper vs. textbook, style, cf. Fig. 1a vs. b), but identical in their lesson plan, learning content and problems to be solved, and taught by the same teacher ("pair classes").

The investigation was conducted in three pairs of school classes in different school types (ST) of secondary level I, two pair-classes in school type 1 and one pair-class in school type 2 (see Table 1). In the three-level system of German secondary education, school type 1 corresponds to the medium educational level ("Realschule", roughly comparable to a British comprehensive school), school type 2 to the highest educational level ("Gymnasium", roughly analogous to a British grammar school on secondary level I, and continuing with secondary level II).[7] These schools were part of a larger cooperation network (Kuhn, 2007, 2008, 2010; Kuhn and Müller, 2007; Kuhn et al., 2008), involving beyond our physics education research group about 40 physics teachers in 15 schools, the latter being actively involved in it by the validation of the instructional material (see below).

Table 1 Study sample and its partition among different school types (CG/TG: control/treatment group).

School year	School type 1	School type 2
school year 1	1 TG (N=16; 8 fem.; 8 m.) 1 CG (N=18; 11 fem.; 7 m.)	1 TG (N=28; 11 fem.; 17 m.) 1 CG (N=26; 10 fem.; 16 m.)
school year 2	1 TG (N=16; 7 fem.; 9 m.) 1 CG (N=18; 12 fem.; 6 m.)	

In total, the tested sample included 122 tenth graders at the age of 15 to 17 with a mean of 16.2 years, 60 students in TG (43% female; 57% male) and 62 students in CG (53% female; 47% male). The age group in the sample is a consequence of German curriculum standards, according to which the topic 'electrical energy' is supposed to be taught in grades 10 of German secondary schools. Before treatment, measures of non-verbal – especially logical – intelligence and reading comprehension as well as a pre-test of motivation (MOT1-PRE) were obtained. In the following three weeks of instruction, the two groups worked on different worksheets containing problems about 'electrical energy' (two physics lessons per week in each group). Problem content, quantity (12 problems per group) and difficulty in the two conditions were identical. After the last worksheet, the students completed a motivation test (MOT2-POST), which was followed by an achievement test. Seven weeks after finishing the following topic, a follow-up motivation test (MOT3-FUP) was conducted to study the long term effect of the treatment[6]. All these measures were obtained by published and standardized instruments, with the exception of the achievement test based on topic related, curriculary valid questions (see section "Materials and Instruments"). The achievement test was also used for grading, in order to keep study related reductions of available teaching time low.

The study design is presented in Table 2.

Instructional material (worksheets)

Worksheets included tasks for practice and knowledge transfer in the pertinent subject matter (energy). Each Worksheet consisted of four tasks with different sub-tasks. The first worksheet dealt with the topics "Electrical Energy". "Electrical Power", "Energy Costs" and with the calculation of these quantities. While the second worksheet calculated the possibilities and limitations of wind energy and atomic energy, the last sheet focused on the discussion of different kinds of energy saving.

In all, students worked on 12 tasks during treatment. The degree of difficulty corresponded to the degree of difficulty of the achievement test. Students worked on the worksheets in groups of two or three. Content and difficulty of the worksheet tasks in the two groups were identical, the

[6]Of course, the same question holds for learning. As positive effects for learning usually are more difficult to achieve than for motivation, the present study tried to establish first whether there is a learning effect at all. Due to classroom restrictions, the question of sustainability had to be restricted to motivation, but ongoing work will extend it to learning (Kuhn, 2010; Kuhn and Müller, in preparation).

[7]See the "TIMSS encyclopedia" (Mullis et al., 2008) for more details concerning the German school system.

Table 2 Study and teaching procedure.

Week	Control group (CG)		Treatment group (TG)	
1	Tests of non-verbal intelligence and reading comprehension Motivation pre-test (MOT1-PRE)		Tests of non-verbal intelligence and reading comprehension Motivation pre-test (MOT1-PRE)	
2	Traditional problems concerning ‘electrical energy’	Worksheet 1	Newspaper-problems concerning ‘electrical energy’	Worksheet 1
3		Worksheet 2		Worksheet 2
4		Worksheet 3		Worksheet 3
5	Immediate motivation test (MOT2-POST) Achievement test		Immediate motivation test (MOT2-POST) Achievement test	
6…13	Traditional education of a new topic			
14	Follow-up motivation test (MOT3-FUP)		Follow-up motivation test (MOT3-FUP)	

Table 3a Competence levels according to PISA (Baumert et al., 2002); problems requiring transfer are level III and above. In the NSP study, problems up to level IV were included.

I: reproduction of simple factual knowledge
II: simple application (mainly using layman concepts)
III: application (using scientific concepts for prediction & explanation)
IV: conceptual and procedural understanding (using elaborated scientific concepts and procedures for prediction & explanation)
(**V**: conceptual and procedural understanding on high level: using scientific models and procedures for differentiated prediction, explanation and analysis)

Table 3b Pisa competence levels of the different items in the achievement post-test and corresponding rating consistency.

Problem-no.	PISA competence level	Cohen's Kappa κ
1	I	0.89
2	II	0.84
3	III	0.82
4	III	0.82
5	IV	0.78

NSP in the TG differed only in the presentation format of the basis text from the tasks in the CG (language style, layout, see Fig. 1).

Finally, the curricular validity of the work sheets was established within the above-mentioned physics education cooperation network; only worksheets with satisfying inter-rater agreement (as measured by Cohen's Kappa (κ_C; Cohen, 1960; Landis and Koch, 1977) were retained (κ_C=0.74-0.91; Kuhn, 2010). For the learning and assessment problems, see the corresponding section below.

Instruments

Motivation measures

Repeated measures of motivation were conducted with an instrument well established in the in the literature on science motivation (adapted from Hoffmann et al., 1997; total Cronbach's α=0.89) with the following subscales: intrinsic motivation (IM; twelve items; Cronbach's α=0.74), classroom climate (CC; ten items; Cronbach's α=0.75) and self-concept (SC; seven items; Cronbach's α=0.82). Individual test scores (‘degree of motivation’ in each subscale) were calculated as the percentage relative to the maximum degree of agreement. The measurement was repeated by the same instrument before, immediately after and seven weeks after treatment (pre/post/follow-up test, MOT1-PRE, MOT2-POST, MOT3-FUP).

Problem difficulty levels and achievement measures

Problems (questions), both for learning worksheets and assessment were discussed and selected according to curricular validity within the physics education network. Competence levels associated with the problems were then operationalized according to the PISA levels (see Table 3a and Baumert et al., 2002). Moreover, these levels were assessed by an expert rating (again with the participating group, other physics teachers and physics education lecturers). Only items with satisfactory rating consistency of curricular validity and level were retained (as measured by κ_C, see Table 3b).

Achievement after treatment (referring to the subject matter electrical energy) was tested with a written test encompassing five different problems, with difficulties similar to those of the worksheets of the training period (see below). Three of these five problems (3, 4, 5) corresponded to the PISA competence levels (PCL) III and IV, involving transfer (application as well as conceptual and procedural scientific understanding used for prediction & explanation), the others to the level I and II (see Table 3b). The format of the problems in the achievement test was conventional for both groups (i.e. *not* newspaper problems), both for reasons of fairness towards the CG (as the test was also used for grading, see “study and teaching procedure” above) and of avoiding bias towards TG. For the same reasons, no items concerning critical reading/thinking were included at this stage of the study. As the content of this intervention (subject matter "energy") had not been executed in one of the lessons or school years before this study, it was completely new and unknown for the students. So we did the intervention without an achievement pre-test. Instead of this prior achievement in physics was assessed as average grade level (average marks in written physics tests) of each student in first six months of the running school term (before the intervention) and was included as an important covariate (see below) to adjust the achievement measures to the students' prior knowledge in physics.

Covariates

Prior achievement in physics was assessed as average grade level (average marks[8] in written physics tests) of each student in first six months of the running school term (before the intervention). Reading comprehension and non-verbal intelligence were assessed by standardized measures and taken into account as covariates, too.[9] The instrument for reading comprehension (Lang et al., 2004) was developed as part of a test battery for career counseling for pupils of the age group of secondary level I (which is considered here), with the reading items focussing on active processing (12 items), and validated within a large sample (N=2556, Cronbach's $\alpha=0.84$). The instrument of non-verbal intelligence (Kornmann and Horn, 2001) was developed as part of a educational screening/counseling battery, with items based on the Figure Reasoning Test (FRT) (25 items) and also validated within a large sample (N=4319, Cronbach's $\alpha=0.81$).

Together with gender, these measures allowed to control for and analyze possible influences of learner features on the effects of the intervention. Moreover, School Type (ST) was included as covariate, due to the general educational level coming along with it.

Methods of analysis

According to the variable plan and the quasi-experimental design described above, ANOVA and ANCOVA were applied as relevant methods (using SPSS in version 22).

Motivation and achievement in physics served as dependent variables, while group membership, school type and gender served as independent variables as well as non-verbal intelligence, reading comprehension and pre-test physics achievement served as covariates.

The reported measure of effect size is omega squared (ω^2), i.e. the population estimate of (total) explained variance, with the usual size categorization (see Cohen,1988: small effects: $0.01<\omega^2<0.06$; medium effects: $0.06\leq\omega^2<0.14$; large effects: $0.14\leq\omega^2$).

[8]The German school grading system has marks from 1 ('very good' resp. A) up to 6 ('insufficient' resp. F).

[9]In this study the test results on reading comprehension and non-verbal intelligence were measured as percentage of correct responses relative to the maximum possible value.

Results

Pre-test comparisons

A 2×2-analysis of variance (ANOVA) was carried out using 'prior achievement level in physics', 'non-verbal intelligence' and 'reading comprehension' as dependent variables and group membership and school type as independent variables (descriptive data: see Table 4). Whereas the groups did not differ in any pre-test variables, the factor 'school type' had a significant but small influence on non-verbal intelligence ($F(1,\ 118)=5.6$; $p<0.05$; $\omega^2=0.04$) and – much stronger – on reading comprehension ($F(1,\ 118)=20.6$; $p<0.01$; $\omega^2=0.14$) before the intervention. This fact was not surprising: because education level in school type 2 is generally significantly more demanding (see PISA-Konsortium Deutschland, 2008), students in this school type are strongly expected to have higher reading comprehension and non-verbal intelligence. For this reason, the covariates in question had to be taken into account.

Furthermore, there was a small, but significant interaction of group membership and school type for motivation (total: $F(1,\ 118)=6.8$; $p<0.05$; $\omega^2=0.05$; "classroom climate" (CC): $F(1,\ 118)=4.8$; $p<0.05$; $\omega^2=0.04$; and "self-concept" (SC): $F(1,\ 118)=6.3$; $p<0.05$; $\omega^2=0.06$). In school type (ST) 1, measures of classroom climate (CC), self-concept (SC) and motivation in total were higher in the TG than in the CG. In contrast, the same measures were lower in the TG than in the CG in ST 2 (see Table 4).

Achievement: post-test analysis

After treatment subject specific physics achievement was tested with the same instrument in both groups. A $2\times2\times2$-analysis of covariance (ANCOVA) was carried out with achievement as dependent variable (post-test according to the section "Problem difficulty levels and achievement measures", analyzed both in total and for its sub-tasks according to different PISA-competence levels separately), group membership, school type and gender as independent variables as well as non-verbal intelligence, reading comprehension and prior achievement in physics as covariates (see Table 5 for descriptive data).

The analysis (see Table 6) shows that group membership had significant and medium size (Problem 2: $F(1,\ 111)=12.5$; $p<0.01$; $\omega^2=0.10$) up to large effects (Problem 3: $F(1,\ 111)=22.5$; $p<0.01$; $\omega^2=0.16$; Problem 5: $F(1,\ 111)=36.0$; $p<0.01$; $\omega^2=0.23$) on the achievement measures (total: $F(1,\ 111)=29.3$; $p<0.01$; $\omega^2=0.20$). Note that the largest values were obtained for problems with competence levels above III (see Table 3a), which, according to their definition involve transfer of knowledge (Baumert et al., 2002).

Prior achievement in physics had significant but small influence on Problem 2 ($F(1,\ 111)=6.6$; $p<0.05$; $\omega^2=0.05$) and Problem 4 ($F(1,\ 111)=4.6$; $p<0.05$; $\omega^2=0.04$), a large influence on Problem 3 ($F(1,\ 111)=29.8$; $p<0.01$; $\omega^2=0.32$) and a medium size effect in total ($F(1,\ 111)=19.5$; $p<0.01$; $\omega^2=0.12$).

For school type, gender and the remaining covariates neither any main effect nor any interaction with group membership were found to be significant.

Table 4 Pre-treatment descriptive data (means (*M*) and standard deviations (SD)) in the different intervention groups and school types.

	Total (*N*=122)		ST1			ST2		
	CG (*N*=62) *M* (*SD*)	TG (*N*=60) *M* (*SD*)	CG (*N*=36) *M* (*SD*)	TG (*N*=32) *M* (*SD*)	Total (*N*=68) *M* (*SD*)	CG (*N*=26) *M* (*SD*)	TG (*N*=28) *M* (*SD*)	Total (*N*=54) *M* (*SD*)
Covariates								
Non-verbal intelligence	79.2 (12.1)	76.9 (11.8)	76.4 (12.1)	74.9 (10.5)	75.8 (11.3)	82.9 (13.4)	79.0 (12.9)	80.9 (13.3)
Reading comprehension	75.1 (17.3)	71.3 (17.1)	68.8 (16.8)	66.7 (19.1)	67.8 (17.8)	85.3 (12.9)	76.5 (12.8)	80.7 (13.5)
Prior achievement in physics	73.9 (14.6)	73.0 (12.3)	73.1 (10.7)	71.9 (9.6)	72.5 (10.1)	74.9 (18.9)	74.3 (14.8)	74.6 (16.8)
Pre-test-data of motivation								
MOT1-PRE, total	51.8 (12.2)	53.7 (12.9)	50.1 (9.5)	57.3 (11.9)	53.5 (11.2)	54.2 (15.2)	49.7 (12.9)	51.9 (14.1)
CC	57.2 (13.4)	57.8 (12.9)	54.6 (11.9)	59.7 (12.3)	56.9 (12.2)	60.8 (14.7)	55.6 (13.4)	58.1 (14.2)
SC	54.5 (14.7)	54.4 (14.7)	52.4 (13.1)	58.1(10.1)	55.1 (12.0)	57.6 (16.4)	50.1 (17.9)	53.7 (17.4)
IM	47.7 (13.1)	50.4 (13.5)	46.9 (10.9)	53.7 (14.4)	50.1 (13.0)	48.8 (15.8)	46.5 (11.3)	47.6 (13.6)

Note: ST1/2: lower/higher educational level ("Realschule/Gymnasium", see Table 1 and sect. "Material and methods"), CC=classroom climate, SC=self-concept, and IM=intrinsic motivation.

Table 5 Post-test descriptive data (means (*M*) and standard deviations (SD)) of achievement in the different groups and school types.

	Total (*N*=122)		ST1			ST2		
	CG (*N*=62) *M* (*SD*)	TG (*N*=60) *M* (*SD*)	CG (*N*=36) *M* (*SD*)	TG (*N*=32) *M* (*SD*)	Total (*N*=68) *M* (*SD*)	CG (*N*=26) *M* (*SD*)	TG (*N*=28) *M* (*SD*)	Total (*N*=54) *M* (*SD*)
T1 (PCL I)	86.5 (21.8)	91.9 (15.8)	82.8 (25.8)	93.2 (13.9)	87.7 (21.5)	91.7 (13.5)	90.5 (17.8)	91.1 (15.8)
T2 (PCL II)	75.0 (30.7)	90.6 (20.0)	66.7 (33.4)	92.6 (14.5)	78.8 (31.7)	86.5 (22.3)	88.4 (24.9)	87.5 (23.5)
T3[aa] (PCL III)	61.1 (39.1)	82.4 (29.1)	61.1 (39.1)	82.4 (29.1)	71.1 (36.1)	-	-	-
T4 (PCL III)	61.3 (40.0)	75.2 (38.1)	50.7 (38.8)	76.2 (35.9)	62.7 (39.4)	75.9 (37.6)	74.1 (41.1)	75.0 (39.1)
T5 (PCL IV)	43.3 (24.7)	68.1 (24.6)	42.4 (23.6)	66.0 (24.5)	53.5 (26.6)	44.5 (26.6)	70.5 (24.9)	57.9 (28.7)
Total	61.1 (20.8)	78.6 (18.7)	56.4 (21.9)	78.8 (17.6)	66.9 (22.8)	67.5 (15.7)	78.3 (20.2)	73.1 (18.8)

Note: ST1/2: lower/higher educational level ("Realschule/Gymnasium", see Table 1 and sect. "Material and methods"), T*x*=sub-item *x* of the achievement test; PCL=PISA competence level.

[a]Item 3 in school type 2 was not taken into account due to incompatible instruction (teacher facilitated solution by giving additional information).

Motivation: within and between subjects-results in the course of time

Motivation was analyzed in a repeated measures design and ANCOVA with treatment groups (group membership (GM) vs. school type (ST) and gender (GR)) as between subjects factors and time of measurement (pre-, post- and follow-up-test; MOT1-PRE vs. MOT2-POST vs. MOT3-FUP) as a within subject factor. Non-verbal intelligence, reading comprehension and prior achievement in physics were used as covariates (see Tables 4 and 7 for descriptive data on MOT1-PRE and MOT2-POST as well as MOT3-FUP, respectively).

Between subject effects (Table 8a): significant and – without exception – large main effects of treatment group were found for overall motivation and all its subscales (classroom climate CC: $F(1, 111)=119.6$; $p<0.01$; $\omega^2=0.45$; self-concept SC: $F(1, 111)=109.8$; $p<0.01$; $\omega^2=0.48$; intrinsic motivation in general IM: $F(1, 111)=92.2$; $p<0.01$; $\omega^2=0.44$; total: $F(1, 111)=125.7$; $p<0.01$; $\omega^2=0.52$).

Significant but small up to medium sized effects were obtained for interactions of group membership with school type (GM × ST; CC: $F(1, 111)=7.4$; $p<0.05$; $\omega^2=0.06$; total: $F(1, 111)=5.8$; $p<0.05$; $\omega^2=0.04$) and with gender (GM × GR; total: $F(1, 111)=4.9$; $p<0.01$; $\omega^2=0.04$) for some subscales and in total measurement of motivation. Also the interaction of group membership with school type and gender became significant but only with small up to medium effects on two of three subscales and on total motivation measurement (GM × ST × GR; CC: $F(1, 111)=7.9$; $p<0.05$; $\omega^2=0.07$; IM: $F(1, 111)=10.3$; $p<0.01$; $\omega^2=0.08$; total: $F(1, 111)=6.0$; $p<0.05$; $\omega^2=0.05$).

Table 6 Post-test comparisons and interactions of achievement between group, school type and covariates: *F*-values (ANCOVA) and effect sizes (ω^2).

	df	Problem 1 (PCL I) F (ω^2)	Problem 2 (PCL II) F (ω^2)	Problem 3 (PCL III) F (ω^2)	Problem 4 (PCL III) F (ω^2)	Problem 5 (PCL IV) F (ω^2)	Total F (ω^2)
Main effects							
Group membership (GM)	1	2.9 (0.02)	12.5** (0.10)	12.8** (0.16)	2.9 (0.02)	36.0** (0.23)	29.3** (0.20)
School type (ST)	1	0.0 (0.00)	0.4 (0.00)	-[a]	2.1 (0.02)	0.1 (0.00)	0.2 (0.00)
Gender (GR)	1	0.3 (0.00)	1.2 (0.01)	-[a]	1.9 (0.02)	0.4 (0.00)	0.0 (0.00)
GM × ST	1	2.3 (0.02)	4.3 (0.05)	-[a]	4.0 (0.03)	0.3 (0.00)	2.4 (0.02)
GM × GR	1	2.0 (0.02)	2.4 (0.02)	-[a]	1.3 (0.02)	2.8 (0.03)	3.0 (0.03)
ST × GR	1	3.1 (0.03)	2.5 (0.02)	-[a]	1.9 (0.02)	3.7 (0.03)	2.9 (0.03)
GM × ST × GR	1	0.3 (0.00)	0.5 (0.00)	-[a]	1.1 (0.01)	0.0 (0.00)	0.2 (0.00)
Covariates							
Prior achievement	1	3.3 (0.03)	6.6* (0.05)	29.8** (0.32)	4.6* (0.04)	1.3 (0.01)	19.5** (0.12)
Non-verbal intelligence	1	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	0.1 (0.00)	1.5 (0.01)	0.5 (0.00)
Reading comprehension	1	2.3 (0.02)	3.6 (0.03)	1.7 (0.02)	0.0 (0.00)	3.1 (0.02)	3.1 (0.02)
Error	111						

Note: PCL=PISA competence level.
*$p<0.05$.
**$p<0.01$.
[a]Item 3 in school type 2 was not taken into account due to incompatible instruction (teacher facilitated solution by giving additional information).

Table 7 Repeated measures descriptive data (means (*M*) and standard deviations (SD)) of motivation in the different intervention groups and school types for post- and follow-up-test, respectively (MOT2-POST, MOT3-FUP; for MOT1-PRE see Table 4).

	Total ($N=122$)		ST1			ST2		
	CG ($N=62$) *M* (*SD*)	TG ($N=60$) *M* (*SD*)	CG ($N=36$) *M* (*SD*)	TG ($N=32$) *M* (*SD*)	Total ($N=68$) *M* (*SD*)	CG ($N=26$) *M* (*SD*)	TG ($N=28$) *M* (*SD*)	Total ($N=54$) *M* (*SD*)
Motivation post-test (MOT2-POST)								
CC	40.4 (9.7)	71.6 (11.6)	42.3 (11.4)	71.9 (12.3)	56.2 (18.9)	37.6 (5.7)	71.32 (10.9)	55.1 (19.1)
SC	38.7 (11.7)	64.7 (10.4)	41.9 (12.7)	64.9 (10.4)	52.8 (16.3)	34.1 (8.5)	64.42 (10.6)	49.8 (18.1)
IM	36.0 (8.6)	63.1 (11.9)	38.9 (8.5)	64.7 (12.9)	51.1 (16.8)	31.9 (7.0)	61.39 (10.7)	47.2 (17.4)
Total	38.5 (8.8)	65.9 (10.2)	41.4 (9.5)	67.1 (11.0)	53.5 (16.5)	34.6 (5.9)	64.61 (9.3)	50.1 (17.1)
Motivation follow-up test (MOT3-FUP)								
CC	48.7 (14.6)	69.3 (12.6)	42.0 (8.7)	71.4 (14.2)	55.8 (18.7)	57.9 (16.1)	66.9 (10.4)	62.6 (14.0)
SC	41.8 (13.6)	63.6 (13.4)	36.3 (8.3)	62.9 (14.2)	48.8 (17.5)	49.4 (15.8)	64.6 (12.5)	57.3 (15.9)
IM	40.9 (13.1)	61.2 (14.0)	37.8 (9.6)	63.8 (16.5)	50.0 (18.6)	45.3 (16.0)	58.3 (10.0)	52.1 (14.7)
Total	43.2 (12.5)	64.1 (12.1)	38.3 (7.6)	65.8 (14.1)	51.2 (17.7)	49.9 (14.9)	62.2 (9.2)	56.3 (13.6)

Note: ST1/2: lower/higher educational level ("Realschule/Gymnasium", see Table 1 and sect. "Material and methods"), CC=classroom climate, SC=self-concept, and IM=intrinsic motivation in general.

As for covariate influences, a significant medium resp. large influence of 'prior achievement in physics' on two of three subscales of motivation was obtained, viz. classroom climate (CC: $F(1, 111)=9.7$; $p<0.05$; $\omega^2=0.08$;) resp. self-concept (SC: $F(1, 111)=27.1$; $p<0.01$; $\omega^2=0.19$;), as well as a significant medium effect on total motivation ($F(1, 111)=9.9$; $p<0.05$; $\omega^2=0.08$;).

No significant main effect of school type and of gender neither on motivation in total nor on any of its subscales was found. The same holds for the rest of the covariates.

Table 8a Time course (repeated measures) comparisons and interactions of motivation between group, school type and covariates: *F*-values (ANCOVA) and effect sizes (ω^2) for between subject effects.

	df	CC F (ω^2)	SC F (ω^2)	IM F (ω^2)	Total F (ω^2)
Between subjects					
Main effects					
Group membership (GM)	1	119.6** (0.45)	109.8** (0.48)	92.2** (0.44)	125.7** (0.52)
School type (ST)	1	0.0 (0.00)	0.3 (0.00)	3.2 (0.03)	1.7 (0.01)
Gender (GR)	1	4.0 (0.03)	3.9 (0.03)	4.2 (0.03)	4.1 (0.03)
GM × ST	1	7.4* (0.06)	3.1 (0.03)	2.7 (0.02)	5.8* (0.04)
GM × GR	1	3.1 (0.03)	2.9 (0.03)	4.0 (0.03)	4.9* (0.04)
ST × GR	1	0.4 (0.00)	0.5 (0.00)	0.1 (0.00)	0.4 (0.00)
GM × ST × GR	1	7.9* (0.07)	3.1 (0.03)	10.3** (0.08)	6.0* (0.05)
Covariate					
Prior achievement in physics (PA)	1	9.7* (0.08)	27.1** (0.19)	1.7 (0.01)	9.9* (0.08)
Non-verbal intelligence (NV)	1	1.6 (0.01)	0.2 (0.02)	2.6 (0.02)	2.8 (0.02)
Reading comprehension (RC)	1	0.1 (0.00)	0.4 (0.00)	0.1 (0.00)	0.8 (0.00)
Error	111				

Note: CC=classroom climate, SC=self-concept, IM=intrinsic motivation in general, PA=prior achievement in physics, NV=non-verbal intelligence, and RC=reading comprehension.
*$p<0.05$.
**$p<0.01$.

Table 8b Time course (repeated measures) comparisons and interactions of motivation between group, school type and covariates: *F*-values (ANCOVA) and effect sizes (ω^2) for within subject effects.

	df	CC F (ω^2)	SC F (ω^2)	IM F (ω^2)	Total F (ω^2)
Within subjects					
Main effects					
Time course (TC)	2	0.3 (0.00)	1.2 (0.10)	0.2 (0.00)	0.2 (0.00)
TC × GM	2	79.1** (0.40)	59.2** (0.34)	57.1** (0.33)	75.8** (0.39)
TC × ST	2	6.1* (0.05)	11.2** (0.09)	3.0 (0.02)	8.1** (0.06)
TC × GR	2	0.0 (0.00)	1.6 (0.01)	1.3 (0.01)	0.1 (0.00)
TC × GM × ST	2	12.7** (0.09)	9.8** (0.07)	6.9* (0.05)	11.9** (0.09)
TC × GM × GR	2	0.3 (0.00)	0.9 (0.01)	0.8 (0.01)	0.7 (0.01)
TC × ST × GR	2	0.2 (0.00)	0.5 (0.00)	0.5 (0.00)	0.5 (0.00)
TC × GM × ST × GR	2	6.3* (0.5)	4.0* (0.4)	3.6* (0.04)	5.9* (0.05)
Covariate					
TC × PA	2	2.1 (0.02)	0.7 (0.01)	3.5 (0.03)	2.3 (0.02)
TC × NV	2	1.5 (0.01)	0.1 (0.00)	0.6 (0.01)	0.5 (0.00)
TC × RC	2	0.7 (0.01)	0.9 (0.01)	0.0 (0.00)	0.5 (0.01)
Error	222				

Note: CC=classroom climate, SC=self-concept, IM=intrinsic motivation in general, GM=group membership, ST=school type, and GR=gender.
*$p<0.05$.
**$p<0.01$.

Within subject effects (Table 8b): both total motivation and all its subscales showed significant and strong interaction of their temporal development (or time course, TC) with group membership (CC: $F(2, 222)=79.1$; $p<0.01$; $\omega^2=0.40$; SC: $F(2, 222)=59.2$; $p<0.01$; $\omega^2=0.34$; IM: $F(2, 222)=57.1$; $p<0.01$; $\omega^2=0.33$; total: $F(2, 222)=75.8$; $p<0.01$; $\omega^2=0.39$). Significant, but only small up to medium main effects on motivation in total and on two of three subscales were found for the interaction "time course vs. school type" (CC: $F(2, 222)=6.1$; $p<0.05$; $\omega^2=0.05$; SC: F

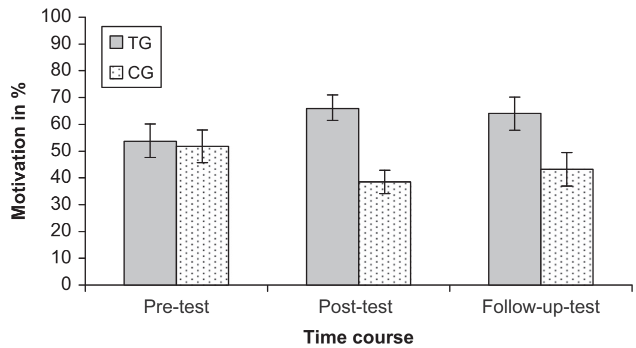


Fig. 2 Time course of motivation (total) in the NSP group and control group (TG and CG, respectively).

(2, 222)=11.2; $p<0.01$; $\omega^2=0.09$; total: $F(2, 222)=8.1$; $p<0.05$; $\omega^2=0.06$).

Finally, significant small up to medium size effects were found for the threefold TC × GM × ST interaction as well as the fourfold TC × GM × ST × GR interaction on each sub-dimension and motivation in total (TC × GM × ST; CC: $F(2, 222)=12.7$; $p<0.01$; $\omega^2=0.09$; SC: $F(2, 222)=9.8$; $p<0.01$; $\omega^2=0.07$; IM: $F(2, 222)=3.6$; $p<0.05$; $\omega^2=0.04$; total: $F(2, 222)=5.9$; $p<0.05$; $\omega^2=0.05$; for TC × GM × ST × GR interaction: see Table 8b).

Thus motivation developed differently in time for different treatment groups and school types. in particular, the strongest significant interaction characterized by large effects on each subscales and motivation in total was found for the TC × GM interaction. As inspection of the time course (Fig. 2) clearly shows, the TC × GM interaction obviously is due to large differences between treatment and control group in development from before to after intervention (and *not* a possible difference afterwards, i.e. from post to follow-up test). Thus, the results of the temporal development of motivation within subjects is consistent with the between subjects main effect of group membership on motivation. The influences of the threefold and the fourfold factor-interaction on each sub-dimension and motivation in total mean, that the positive motivation development of TG compared to CG was different for both school types ST1/2 considered (as clearly visible when comparing Figs. 3 and 4). This might be due to chance, but a plausible explanation is as follows: in school type 1 (“Realschule”) grade 10, where the instruction took part, is the last year in school; a general drop of motivation towards the very end of the schooling period is a well-known experience of many partitioners supported by the data (Fig. 3, CG curve). This factor does not exist for school type 2 (“Gymnasium”, Fig. 4), hence the difference found.

No influence of gender or of any interaction of gender and other factors neither on motivation in total nor on subscales could be discovered. The holds true for the influence of all other learner covariates considered.

Discussion

With regard to the hypotheses and research questions of this study, the results can be interpreted as follows:

Motivation differences between groups: the motivation in the treatment group (learning with newspaper story

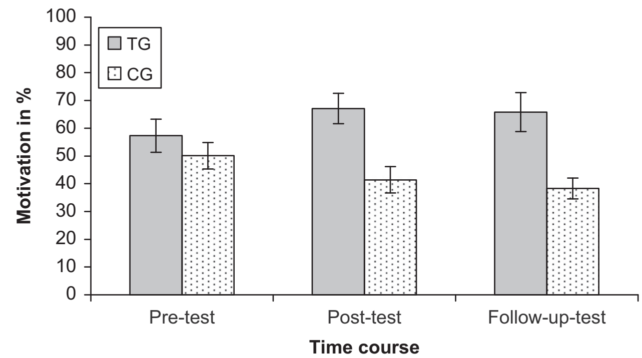


Fig. 3 Time course of motivation (total) in the NSP group and control group as analyzed for school type 1 (lower educational level; Realschule; TG and CG, respectively).

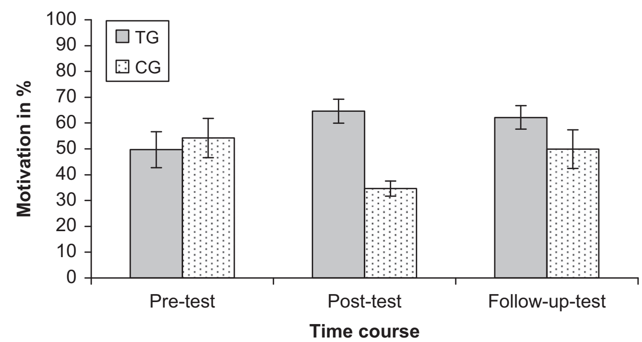


Fig. 4 Time course of motivation (total) in the NSP group and control group as analyzed for school type 2 (higher educational level; Gymnasium; TG and CG, respectively).

problems) led to a very considerable improvement of motivation compared to a control group (learning with problems without narrative context, but with identical content). Effects were throughout statistically very highly significant and of very large effect sizes for all motivations subscales (CC: $\omega^2=0.45$; SC: $\omega^2=0.48$; MI: $\omega^2=0.44$; total: $\omega^2=0.52$; these are the values averaged over post and follow up test). Note that the largest effect was found for self-concept, a motivation component where NSP can be theoretically expected to be particularly helpful. These findings support hypotheses 1 and 3.

Achievement: learning in the treatment group was considerably enhanced when compared to the control group. Effects are statistically at least very significant, and effect sizes generally large (total: $\omega^2=0.20$). Thus, NSP lead to improved learning with an effect size of considerable practical importance. This holds in particular also for more demanding items, assessing (among other) students' transfer abilities. These findings support hypotheses 2 and 4.

Time course of motivation/temporal stability of effects (repeated measures; see Fig. 2): In the treatment group, motivation increased most significantly (compared to initial state) and lasted for several months, whereas the motivation in the control group decreased after treatment and stayed low for the following months. This different temporal development of motivation in TG compared to CG was most significant and of very large effect size (CC: $\omega^2=0.40$; SC: $\omega^2=0.34$; IM: $\omega^2=0.33$; total: $\omega^2=0.39$). Thus, the narrative contexts of NSP lead to an improved motivation when compared to the initial state, showing (at least) medium-term stability. These results answer the research question 6 (where no hypothesis was made) in the positive sense.

Influences of prior knowledge in physics, non-verbal intelligence, reading comprehension, gender and school-type: while prior knowledge had a significant (positive) influence on the achievement and motivation (as expected), there was no interaction with treatment group. Thus, the NSP approach does not privilege students with higher prior knowledge more than traditional instruction. No influences of non-verbal intelligence and reading comprehension were found on any component of motivation nor on achievement (both in total or for any of its sub-items), and the same holds true for gender main effects or interaction with group membership. Moreover, overall size and stability of motivation could be replicated in each school type, showing no interaction with motivation, (except for a small interaction effect with the time course of motivation, favouring school type 2; see end of the results section "Motivation: within and between subjects-results in the course of time" for a possible explanation). So the effects do not depend on the general educational level coming along with school type, either. These findings support our hypotheses 5: the beneficial effects for both motivation and learning do not (or weakly) depend on learner characteristics (such as academic level) nor on classroom/school characteristics (such as school type and others). In particular, the gender independence stated by Fensham (2009) for story contexts could be replicated.

In summary, the hypotheses put forward for newspaper story problems as specific form of CBSE are supported by the data: the intervention led to both improved motivation in general, and self-concept in particular, as well as to improved learning in general, and transfer in particular with most effect sizes being large (medium in some cases). As for the remaining research questions, motivation gains lasted at least for several months (sustainability), and the beneficial effects result held regardless of various class and learner characteristics, such as general education/school level, gender, various aspects of ability, and others) (robustness).

Furthermore, note a number of methodological exhortations put forward by recent reviews (Bennett et al., 2007; Taasoobshirazi and Carr, 2008), and being relevant for the present research. First, we took several measures to minimize teacher influence: treatment and control pair classes were taught by the same teachers. These had not participated in the development of the instructional material, in order to minimize a possible identification with the new approach. Furthermore, the active learning phase proper was independent work by the pupils, covering 3/4 of available instruction time, where the teachers did almost not intervene at all (or to a negligible extent). While these measures do not allow for a complete control of teacher influences, they represent a step forward in the sense of the above-mentioned exhortations, and are compatible with the practical limitations of the real-life classroom teaching the study was embedded in. Second, the importance of considering possible influences of student's characteristics (e.g. gender, ability) and prerequisites (e.g. reading comprehension) has been repeatedly stressed (for CBSE: see e.g. Bennett et al., 2007; more generally: Seidel and Shavelson, 2007). Moreover, Taasoobshirazi and Carr (2008) conclude their review on context based physics education stating frequent methodological problems such as lack of pretests, control groups, and of measures of learning (even though being a central goal of context based approaches), leaving the number of studies in compliance with these requirements very low, and consequently with an urgent need of more work of this kind. Following closely these exhortations, we believe to have contributed to the kind of research required by these reviews, by carrying out a quasi-experimental design including motivation and learning measures, accounting for the pre-instruction state of affairs and a number of other pertinent classroom and student influences, but still remaining in accordance with practical classroom requirements as well.

Conclusions and outlook

The present contribution combines "good practice" reports on the promising use of newspaper story problems in science (and mathematics) education with empirical research, based on a theoretical background on context based learning with an emphasis on narrative contexts on the one hand, and design principles inspired by anchored instruction on the other.

This approach was investigated in a quasi-experimental study (on energy and energy transformations in German 10th grade classes) with a number of control measures: same teacher in treatment and control group, identical learning sequence and learning tasks (up to their fundamental format, viz. newspaper vs. conventional format), consideration of potentially influential cofactors and covariates. Instructional material and classroom setting (time course, form of student activity) in both groups were tested for curricular validity in a physics education cooperation network, involving more than 40 teachers from various backgrounds of secondary level in the study country.

Under the double constraints of classroom practice and educational research, the findings of the study contribute to the questions raised e.g. in the research synthesis of Bennett et al. (2007): as main or general effects, newspaper story problems led to improved motivation and learning, including transfer, with effect sizes between medium and large (motivation (total): $\omega^2=0.52$; learning/achievement (total): $\omega^2=0.20$; transfer (average): $\omega^2=0.14$). As for possible differential effects, such as possibly different outcomes for girls and boys or students of different ability, no (or weak) influences of this kind could be found. This means in particular that the absolute level of understanding attained by the low ability group being less than that of high ability group, as might be expected, their *gain* when learning with NSP (instead with conventional problems) turned out to be the same. Moreover, an interesting finding, viz. the gender neutrality found in PISA and supposedly attributed to story contexts (Fensham, 2009) could be replicated. Another issue of particular interest for science literacy (almost by its definition) is transfer of learning;

here too, considerable benefits through learning with newspaper story problems have been found.

In view of much more far reaching changes of the teaching script by various forms of CBSE (as described in Bennett et al. (2007), such as the STS approach of the Iowa project (Yager and Weld, 1999) or Anchored Instruction (CTGV, 1992; CTGV, 1997) possible doubts regarding more restricted approaches such as NSP may arise, and concerns about limitations of its benefits are of particular interest, such as small size and short temporal duration of possible effects as well as too narrow a restriction of student groups profiting from them. At least for these three kinds of limitations (for others, see below), the results of the study are encouraging: first, effects sizes for motivation and learning are of practical importance (see above); second, these benefits are robust with respect to a number of a priori important influences (non-verbal intelligence, reading comprehension and school type, i.e. general education level); third, motivation is stable at least in medium term (four months).

To our mind, the NSP approach with its double roots in context based science learning and design principles inspired by Anchored Instruction has shown its raison d´être in that it shows useful benefits, and it does so with a classroom setting and learning media which are inexpensive in time and money, flexible and easy to modify, thus meeting important demands of practitioners. We will now turn to some implications and perspectives for both future research and classroom practice.

Implications for future research

Guided by the above-mentioned exhortations (Bennett et al., 2007; Seidel and Shavelson, 2007; Taasoobshirazi and Carr, 2008), the following research questions should be further examined in the theoretical and methodological framework of the present study:

1. To investigate further generalizability and flexibility as essential features of classroom implementation, research will be expanded to other populations (e.g. age groups, school types and educational levels) and subject matters (in physics and other sciences). In particular, the applicability of the approach for students with low educational level deserves further attention. In the present study, medium academic level schools within the three-level system of German secondary education were included, and no influence of general or disciplinary level (regarded as covariates) was found. But there is a 3rd school type ("Hauptschule") with generally lowest academic level and socio-cultural background, and where the applicability of the approach will be investigated, too.
2. In addition to Newspaper story problems, different types of context-oriented learning problems with narrative embedding should be studied (e.g. design variants with and without pictures, advertisements, etc.), combining them also with other ways of establishing contexts (visual, hands-on activity, etc.).

   The didactical flexibility and variability of the NSP-approach is a prerequisite for two further research questions, requiring context based instructional materials and settings which can easily be changed also with respect to the following characteristics:
3. How do different degrees of openness of the newspaper story problems influence achievement and motivation? First attempts of operationalisation of the degree of openness of problems do exist (Baumert, 1996; Mehlhorn and Mehlhorn, 1979) and have been applied to newspaper story problems (Kuhn, 2007, 2008; Kuhn and Müller, 2006, 2007).
4. It is well established that too complex and demanding task requirements reduce learning success or prevent it altogether (cognitive load theory, Sweller et al., 1990). On the other hand, features of situated learning like real life 'complexity' obviously lead to highly demanding tasks. Consequently, further research should try to answer the following research question: how does complexity and cognitive load entailed by various contexts affect motivation and learning?

Moreover, the following issue is of considerable theoretical and practical interest: a factor common to many context based approaches is "authenticity" and relatedness to real life. It is quite current in CBSE to consider "authenticity" as so essential for "context", that the two form a kind of natural unit, such that the combined terms "authentic contexts" often occur almost inseparately (see e.g. in science education Schwartz et al., 2004; Aikenhead, 2006; PISA-Konsortium Deutschland, 2008; in general education Vosniadou, 2001; Herrington and Herrington, 2006; Sawyer, 2009). But it is authenticity *for the learner*, which is the crucial point, i.e. her or his subjective perception, not authenticity for the teacher nor researcher. For a better understanding, which factor might make a particular form of CBSE more successful than another, one thus needs (among other things) an instrument to assess perceived authenticity as manipulation check. Development of such an instrument is under way in our working group (Kuhn and Müller, in preparation), and it will allow to analyze NSP as well as other context based approaches more in detail.

Finally, we turn to the following two types of possible limitations. First, as for educational *characteristics*, future research (and its applications) on important issues such as complexity and openness (see above) might by facilitated by the NSP realization of context based learning and its flexibility; This, in turn, might eventually lead to an improved understanding of macro approaches, rather than representing a limitation. Second, we again consider possible limitations of educational objectives. Beyond those mentioned above, and turning out *not* to occur for NSP (small, narrow, and short term effects only), there is an important limitation, which actually *is* at work in the present study and its results. It is about higher order competences like critical thinking in general and critical reading of science related media reports in particular (Norris and Phillips, 1994, 2003; Millar and Osborne, 1998; Wellington and Osborne, 2001; McClune and Jarman, 2010). The same is true for still more general higher order competences, such as problem solving, awareness of the decisive importance of "science and society" issues, and, eventually, responsible citizenship in the full breadth of sense of scientific literacy. While these educational aims are obviously important, they are not easy to assess. The main objective of the present work was to establish, whether NSP have enough effectiveness to be of practical importance, which seems to us an important issue, too, looking at the generally quite small, zero or even negative effect sizes reported for existing CBSE interventions (Bennett et al., 2007; Taasoobshirazi and Carr, 2008). Given this state of affairs, and the restrictions of the practical classroom conditions (above all, allocated time), it was neither feasible nor appropriate to include assessment of these competences in the present study. Taking it now, however, as starting point, with increased confidence in the basic effectiveness, it is possible and important to go beyond this limitation. Thus, beyond an assessment of perceived authenticity (see above), also assessment for higher-order competences has to be included in future research, drawing on existing work (such as for critical thinking, see e.g. Scriven and Fisher, 1997).

Implications and further perspectives for classroom practice

The entire rationale for the present study is an attempt to bring together the advantages of narrative contexts and of essential design principles of anchored instruction (such as authentic problems and embedded data) with features as availability, practicability and flexibility put forward by teachers as classrooms practitioners of CBSE. The results show, that using the NSP intervention in everyday physics lessons can contribute to an affectively and cognitively stimulating teaching and it does so:

- with an amount of cost and effort well within the usual possibilities of teachers and schools and
- in a broad variety of conditions, both on the individual (in particular, for both genders) and superordinate levels (in particular, when used by different teachers), and easily adaptable to a given classroom situation.

We thus consider it as an encouraging way for teachers of pursuing the CBSE line of thought, combining it with good practice evidence and adaptable it to their own needs.

While this perception by classroom practitioners could be empirically confirmed among the teachers involved in the development and implementation of this study (Kuhn, 2010; Kuhn et al., 2008), the broadening of its range of application to other science education topics (e.g. chemistry) and similar learning media (e.g. problems based on advertisements) is an obvious generalization which was proposed to us by many teachers (see also section above). Several ideas of this kind were already tried out and investigated within the classroom research and development network established since the beginning of the research project (Kuhn, 2005, 2010; Kuhn et al., 2010; Kuhn and Müller, in preparation).

We consider this line of thought as important in order to increase further the applicability and practicability of the NSP approach, combine in with other instructional approaches, and, when pursued further, to modify and broaden also future research directions. Thus, implications and perspectives for classroom practice are in close connection with implications for future investigations (points 1 and 2 above). The same holds for the research agenda concerning openness and complexity, which are obviously also relevant for classroom implementation. This contribution should in no way be read as a pleading for an exclusively NSP-based curriculum, in view of the limitations of the study (such as duration and educational objectives considered), and because teaching and learning live on a variety of methods. But, concluding with Fensham (2009) we feel "that 'Science as a Story' need to become a quite central pedagogy in science teaching", and that newspaper story problems might offer a useful contribution to that purpose.

Conflict of interest

None of the authors have any conflict of interest.

References

Aikenhead, G., 2006. Science Education for Everyday Life: Evidence-Based Practice. Althouse Press, London, Ontario.

Anderson, J.R., 2010. Cognitive Psychology and Its Implications. Worth Publishers, New York.

Armbrust, A., 2001. Physikaufgaben und -informationen aus der Zeitung, Der Mathematisch-Naturwissenschaftliche Unterricht (MNU). 54, 405–409.

Baumert, J., 1996. Technisches Problemlösen im Grundschulalter: Zum Verhältnis von Alltags- und Schulwissen – Eine kulturvergleichende Studie. In: Leschinsky, A. (Ed.), Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule. Beltz Verlag, Basel, Weinheim, pp. 187ff.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., Weiß (Eds.), M., 2001. PISA 2000. Basiskompetenzen von Schülern und Schülerinnen im internationalen Vergleich. Leske+Budrich, Opladen.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann, K.-J., Weiß, M., 2002. PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich: Zusammenfassung der Befunde, In: Deutsches PISA-Konsortium (Eds.), Länder der Bundesrepublik Deutschland im Vergleich. Leske+Budrich, Opladen, 35.

Bennett, J., Lubben, F., Hogarth, S., 2007. Bringing science to life: a synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. Sci. Educ. 91 (3), 347–370.

Blumschein, P., 2003. Eine Metaanalyse zur Effektivität multimedialen Lernens am Beispiel der Anchored Instruction (Unveröffentlichte Dissertation). Albert-Ludwigs Universität, Institut für Erziehungswissenschaften, Freiburg, Germany.

Bottge, B.A., Heinrichs, M., Metha, Z.D., Hung, Y.-H., 2002. Weighing the benefits of anchored instruction for students with disabilities in general education classes. J. Spec. Educ. 35 (4), 186–200.

Brahler, C.J., Peterson, N.S., Johnson, E.C., 1999. Developing on-line learning materials for higher education: an overview of current issues. Educ. Technol. Soc. 2.

Bransford, J.D., Sherwood, R.D., Hasselbring, T.S., Kinzer, C.K., Williams, S.M., 1990. Anchored instruction: why we need it and how technology can help. In: Nix, D., Spiro, R. (Eds.), Cognition, Education and Multimedia. Erlbaum, Hillsdale, NJ, pp. 115–141.

Bransford, J.D., Stein, B.S., 1993. The Ideal Problem Solver. Freeman, New York.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20 (1960), 37–46.
Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum, Hillsdale, NJ.
CTGV, 1990. Anchored instruction and its relationship to situated cognition. Educ. Res. 19 (6), 2-10.
CTGV, 1991. Technology and the design of generative learning environments. Educ. Technol. 31, 34-40.
CTGV, 1992. The Jasper series as an example of anchored instruction: theory, program description and assessment data. Educ. Psychol. 27, 291–315.
CTGV, 1993. Anchored instruction and situated cognition revisited. Educ. Technol. 33 (3), 52–70.
CTGV, 1997. The Jasper Project: Lessons in Curriculum, Instruction, Assessment, and Professional Development. Erlbaum, Hillsdale, NJ.
Fensham, P.J., 2009. Real world contexts in PISA science: implications for context-based science education. J. Res. Sci. Teach. 46, 884–896.
Gardner, G.E., Jones, M.G., Ferzli, M., 2009. Popular media in the biology classroom: viewing popular science skeptically. Am. Biol. Teach. 71 (6), 332-335.
Gilbert, J.K., Bulte, A., Pilot, A., 2011. Concept development and transfer in context-based science education. Int. J. Sci. Educ. 33 (6), 817–837.
Glaser, R., Carson, K, 2005. Chemistry is in the news: taxonomy of authentic news media-based learning activities. Int. J. Sci. Educ. 27 (9), 1083–1098.
Glynn, S., Koballa, T.R., 2005. The contextual teaching and learning approach. In: Yager, R.E. (Ed.), Exemplary Science. Best Practices in Professional Development. National Science Teacher Association Press, Arlington, VA, pp. 75-84.
Hattie, A.C., 2009. Visible Learning. A Synthesis of Over 800 Meta-analyses Relating to Achievement. Routledge, London, New York.
Haupt, P., 2005. Zeitungsberichte als Arbeitsmaterial für den Unterricht – Start einer neuen Rubrik. Naturwissenschaften im Unterricht-Chemie (NiU-Chemie). 11 (55), 46-50.
Herget, W., Scholz, D., 1998. Die etwas andere Aufgabe – aus der Zeitung. Mathematik-Aufgaben Sek. I. Kallmeyersche Verlagsbuchhandlung, Seelze.
Herrington, A., Herrington, J. (Eds.), 2006. Authentic Learning Environments in Higher Education. Information Science Publishing, Hershey, London.
Hoffmann, L., Häußler, P., Peters-Haft, S., 1997. An den Interessen von Mädchen und Jungen orientierter Physikunterricht. Ergebnisse eines BLK-Modellversuches. Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel, Kiel.
Hoots, R.A., 1993. Biology in the news. Am. Biol. Teach. 53 (8), 496–499.
Jarman, R., McClune, B., 2001. Use the news: study of secondary teachers' use of newspapers in the science classroom. J. Biol. Educ. 35 (2), 69–74.
Jarman, R., McClune, B., 2002. A survey of the use of newspapers in science instruction by secondary teachers in Northern Ireland. Int. J. Sci. Educ. 24 (10), 997–1020.
Jarman, R., McClune, B., 2007. Developing Scientific Literacy. Using News Media in the Classroom. Open University Press, Maidenhead.
Kornmann, A., Horn, R., 2001. SSB – Screeningverfahren für Schul- und Bildungsberatung. Teil 2. Swets Test Services GmbH, Frankfurt a. M., Germany.
Kossack, S., 1987. Use the news. J. Read. 30 (6), 552-554.
Kuhn, J., 2005. The modified 'anchored instruction'-approach: anchor-media and 'task-culture' in physics education within the theoretical framework of situated learning. Acta Syst. – IIAS Int. J. V (1), 17–26.
Kuhn, J., 2007. Authentische Aufgaben im Physikunterricht: Nachhaltige Bildung durch Entwicklung von Ankermedien und 'Kultivierung' von Aufgaben. In: Lemmermöhle, D., Rothgangel, M., Bögeholz, S., Hasselhorn, M., Watermann, R. (Eds.), professionell lehren – erfolgreich lernen. Waxmann Verlag, Münster, pp. 251–263.
Kuhn, J., 2008. Optimierung authentischer Aufgaben als Ankermedien im Rahmen einer neuen 'Aufgabenkultur' im Physikunterricht. In: Höttecke, D. (Ed.), Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Essen 2007. LIT-Verlag, Münster, pp. 281–284.
Kuhn, J., 2010. Authentische Aufgaben im theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht. Vieweg+Teubner Verlag, Wiesbaden.
Kuhn, J., Müller, A., 2005a. Ankermedien und 'Aufgabenkultur' im Physikunterricht: Zwei empirische Studien im theoretischen Rahmen des situierten Lernens. In: Nordmeier, V., Oberländer, A. (Eds.), Didaktik der Physik. Beiträge zur Frühjahrstagung der DPG – Berlin 2005 [CD]. Lehmanns Media, Berlin.
Kuhn, J., Müller, A., 2005b. Ein modifizierter 'Anchored Instruction'-Ansatz im Physikunterricht: Ergebnisse einer Pilotstudie. Empir. Pädag. (EP) 19 (3), 281-303.
Kuhn, J., Müller, A., 2006. Authentische Aufgaben – 'Zeitungsaufgaben' als Beispiel zur Umsetzung von Bildungsstandards in Physik. Praxis der Naturwissenschaften – Physik in der Schule (PdN-PhiS) 55 (4), 29–34.
Kuhn, J., Müller, A., 2007. Authentische Aufgaben zur Kompetenzausrichtung: Kriterien zur Gestaltung offener Aufgaben. Praxis der Naturwissenschaften – Physik in der Schule (PdN-PhiS) 56 (6), 38–45.
Kuhn, J., Müller, A., 2014. An instrument for motivation and authenticity in science education (in preparation).
Kuhn, J., Müller, A., Schneider, C., 2008. Das Landauer Programm zur Lehrerbildung in den Naturwissenschaften (LeNa): Standardbezogene Evaluation und Interventionen für eine verbesserte Abstimmung auf dem Prüfstand. Empirische Pädagogik (EP) 22 (3), 305-327.
Kuhn, J., Müller, A., Müller, W., Vogt, P., 2010. Kontextorientierter Physikunterricht: Konzeptionen, Theorien und Forschung zu Motivation und Lernen. Praxis der Naturwissenschaften – Physik in der Schule (PdN-PhiS) 5 (59), 13-25.
Kultusminister der Länder in der Bundesrepublik Deutschland/ Bundesverbands Deutscher Zeitungsverleger (KMK/BDZV), 2006. Gemeinsame Erklärung des Bundesverbandes Deutscher Zeitungsverleger und der Präsidentin der Ständigen Konferenz der Kultusminister der Länder zur Kooperation zwischen Zeitungsverlagen und Schulen: Projekt, SCHmitZ – Schule mit Zeitung. KMK, Berlin. ⟨http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_10_19-PM-SCHmitZ.pdf⟩ (03.2014).
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Lang, D., Mengelkamp, C., Jäger, R.S., 2004. Entwicklung von Testverfahren zur Berufsberatung von Schülern. Empirische Pädagogik (EP) 18 (3), 281ff.
Mandler, J.M., 1984. Stories, Scripts and Scenes: Aspects of Schema Theory. Lawrence Erlbaum, Hillsdale, NJ.
Mandler, J.M., 1987. On the psychological reality of story structure. Discourse Process. J. 10, 1–29.
Mandler, J.M., 2004. The Foundations of Mind: The Origins of Conceptual Thought. Oxford University Press, New York.
Mandler, J.M., Johnson, N.S., 1977. Remembrance of things passed: story structure and recall. Cogn. Psychol. 9, 111-151.
McClune, B., Jarman, R., 2010. Critical reading of science-based news reports: establishing a knowledge, skills and attitudes framework. Int. J. Sci. Educ. 32 (6), 727–752.
Mehlhorn, G., Mehlhorn, H.-G., 1979. Untersuchungen zum schöpferischen Denken bei Schülern, Lehrlingen und Studenten. Volk und Wissen, Berlin.

Millar, R., Osborne, J., 1998. Beyond 2000: Science Education for the Future. King's College, London.
Mullis, I.V.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., Stanco, G.M. (Eds.), 2008. TIMSS 2007 Encyclopedia: A Guide to Mathematics and Science Education Around the World (Volumes 1 and 2). TIMSS & PIRLS International Study Center, Chestnut Hill, MA.
Müller, A., Kuhn, J., Müller, W., Vogt, P., 2010. Modified anchored instruction" im Naturwissenschaftlichen Unterricht: Ein Interventions- und Forschungsprogramm. In: Höttecke, D. (Ed.), Entwicklung naturwissenschaftlichen Denkens – zwischen Phänomen und Systematik. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Dresden 2009. LIT-Verlag, Münster, pp. 149–151.
Newspaper Association of America Foundation (NAAF), 2007. NIE: Newspapers in Education – Lifelong Readers-Driving Civic Engagement. NAAF, Arlington, VA. ⟨http://www.naafoundation.org/docs/Foundation/Research/LR_civic.pdf⟩ (03.2014).
Newspaper Association of America Foundation (NAAF), 2010a. NIE: Newspapers in Education – Common Threads Linking NAA Foundation Research to Today's Young Media Consumers. NAAF, Arlington, VA. ⟨http://www.naafoundation.org/Research/Foundation/NIE/Common-Threads.aspx⟩ (3.2014).
Newspaper Association of America Foundation (NAAF), 2010b. NIE in 2010. NAAF, Arlington, VA.
Newspaper Association of America Foundation (NAAF), 2011. History of NIE. NAAF, Arlington, VA. ⟨http://www.naafoundation.org/About/Programs/NIE/History-Of-NIE.aspx⟩ (03/2014).
Norris, S.P., Phillips, L.M., 1994. Interpreting pragmatic meaning when reading popular reports of science. J. Res. Sci. Teach. 31, 947–967.
Norris, S.P., Phillips, L.M., 2003. How literacy in its fundamental sense is central to scientific literacy. Sci. Educ. 87, 224–240.
OECD, 2006. Assessing Scientific, Reading and Mathematical Literacy. A Framework for PISA 2006. ⟨http://www.oecd.org/dataoecd/63/35/37464175.pdf⟩ (3/2014).
Paulos, J.A., 1995. A Mathematician Reads the Newspaper. Basic Books, New York.
PISA-Konsortium Deutschland (Ed.), 2008. PISA 2006 in Deutschland – Die Kompetenzen der Jugendlichen im dritten Ländervergleich. Waxmann-Verlag, Münster.
Renkl, A., Mandl, H., Gruber, H., 1996. Inert knowledge: analyses and remedies. Educ. Psychol. 31, 115–121.
Rhoades, L., Rhoades, G., 1980. Teaching With Newspapers: The Living Curriculum. Phi Delta Kappa Educational Foundation, Bloomington.
Roberts, D.A., 2007. Scientific literacy/science literacy. In: Abell, S.K., Lederman, N.G. (Eds.), Handbook of Research on Science Education. Routledge, New York, pp. 729–780.
Rumelhart, D.E., 1975. Notes on a schema for stories. In: Bobrow, D.G., Collins, A. (Eds.), Representation and Understanding: Studies in Cognitive Science. Academic Press, New York.
Salkind, N.J. (Ed.), 2008. Encyclopedia of Educational Psychology. Sage Publications, Thousand Oaks.
Sawyer, R. (Ed.), 2009. The Cambridge Handbook of the Learning Sciences. University Press, Cambridge.
Schwartz, R.S., Lederman, N.G., Crawford, B.A., 2004. Developing views of nature of science in an authentic context: an explicit approach to bridging the gap between nature of science and scientific inquiry. Sci. Educ. 88, 610–645.
Scriven, M., Fisher, A., 1997. Critical Thinking: Its Definition and Assessment. Point Reyes Edgepress, Norwich.
Seidel, T., Shavelson, R.J., 2007. Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. Rev. Educ. Res. 77 (4), 454–499.
Shavelson, R.J., Hubner, J.J., Stanton, G.C., 1976. Self-concept: validation of construct interpretations. Rev. Educ. Res. 46, 407–444.
Shymansky, J.A., Kyle, W.C.J., Alport, J.M., 1983. The effects of new science curricula on student performance. J. Res. Sci. Teach. 20 (5), 387–404.
Sweller, J., Chandler, P., Tierney, P., Cooper, M., 1990. Cognitive load as a factor in the structuring of technical material. J. Exp. Psychol.: Gen. 119, 176–192.
Taasoobshirazi, G., Carr, M., 2008. A review and critique of context-based physics instruction and assessment. Educ. Res. Rev. 3 (2), 155–167.
Toby, S., 1997. Chemistry in the public domain: a plethora of misinformation – or, don't believe everything you read in the newspapers!. J. Chem. Educ. 74, 1285–1287.
Uguroglu, M.E., Walberg, H.J., 1979. Motivation and achievement: a quantitative synthesis. Am. Educ. Res. J. 16 (4), 375–389.
Vosniadou, S., 2001. How Children Learn. International Academy of Education, Geneva.
Wellington, J., Osborne, J.F., 2001. Language and Literacy in Science Education. Open University Press, Buckingham.
Wild, E., Hofer, M., Pekrun, R., 2001. Psychologie des Lerners. In: Krapp, A., Weidenmann, B. (Eds.), Pädagogische Psychologie – Ein Lehrbuch. Beltz Psychologie Verlag, Weinheim, pp. 207–270.
Yager, R.E., Weld, J.D., 1999. Scope, sequence and coordination: the Iowa Project, a national reform effort in the USA. Int. J. Sci. Educ. 2/21 (1999), 169–194.
Zabel, J., 2004. Narrative Strukturen beim Lernen der Evolutionstheorie – Das Verständnis evolutiver Prozesse in Lernergeschichten zur Walevolution. In: Vogt, H., Krüger, D., Herget, M., Bögeholz, S. (Eds.), Erkenntnisweg Biologiedidaktik. Sektion Biologiedidatik, Berlin, München, pp. 95-113. ⟨http://www.biologie.fu-berlin.de/didaktik/Erkenntnisweg/2004/index.html⟩ (3.2014).
Zabel, J., 2007. Stories and meaning: what students' narratives reveal about their understanding of the theory of evolution. In: Proceedings of the 2007 Conference of the European Science Education Research Association (ESERA), Malmö, Sweden. ⟨http://na-serv.did.gu.se/ESERA2007/ESERA2007.htm⟩ (3/2014).